

Single and Multiple Instance Learning for Visual Categorisation



Ruo Du

Faculty of Engineering and Information Technology
University of Technology, Sydney

A thesis submitted for the degree of

Doctor of Philosophy

2013

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Author

Ruo Du

This thesis is dedicated to my wife and my parents.

For their endless support and encouragement

Acknowledgements

First and foremost I would like to sincerely thank my principal supervisor, Prof. Xiangjian HE, for his guidance, understanding, patience and help on scholarship to make my PhD experience such a rewarding and exciting journey.

I also want to express my utmost gratitude to my co-supervisor, Dr. Qiang Wu, who spent enormous time and energy on me as a mentor and a friend. His guidance was embedded in every step of my studies.

Special thanks also to Dr. Wenjing Jia, Dr. Min Xu, Prof. Massimo Piccardi and Dr. Richard Xu for their great suggestions, knowledge sharing and invaluable assistance.

I wish to thank my fellow research students of our team, Muhammad Hasan, Man Wong, Chao Zeng, Sheng Wang, Minqi Li, Thomas Tan, Aruna Jamdagni, Ying Wan and Mohammed A AmbuSaid, for their assistance and friendships.

Finally, and most importantly, I would like to thank my wife Ting Zhou. Her faith in me, support, encouragement and quite patience made it possible that I could even continue to pursue my PhD after working in IT for ten years. I thank my parents, Changyou Du and Shenghua Liu, and my parents in law, Ruguo Zhou and Shijie Li, for their help and support as always.

Abstract

Nowadays, huge amounts of visual data, e.g., videos and images, have become widely accessible. Therefore, intelligently categorizing the large and growing collections of data for access convenience has been a central goal for modern computer vision research. In this thesis, we describe several newly-developed approaches for visual categorization upon the single and multiple instance learning cases.

In single-instance learning (SIL), each of the training instances has been labeled. Here, we focus on a challenging task of facial expressions recognition where manually labeling each training instance, i.e., face video, is handy. To get the distinct features of expressions, we propose a novel feature representation, Histogram Variances Face (HVF), which integrates dynamic expression information into a static image being invariant to illumination and in-plane rotation. Through HVFs, the facial expression recognition can be cast as a facial recognition problem. We have applied our approach on the well-known Cohn-Kanade AU-Coded Facial Expression database, and then those extracted HVFs are classified by using facial recognition technology, i.e., Eigenfaces and Support Vector Machines (SVMs). The recognition accuracy is very encouraging. We further propose an extension of HVFs, Hexagonal Histogram Variance Faces (HHVFs), which applies HVFs on a hexagonal structure. Comparing to HVFs, HHVFs not only greatly reduce the computation costs but also improve the recognition accuracy.

In multiple-instance learning (MIL), the training instances are divided into groups and the instances in the same group share only one label. MIL arises from many applications where individually labeling training instances is expensive. In this case, we propose a novel algorithm,

multiple-instance learning with a supervised kernel density estimation (MIL-SKDE), to tackle the labeling ambiguity. Our algorithm extends the twin technologies, kernel density estimation (SKDE) and mean shift, to their supervised versions in which the labels of data points will affect the mode seeking. We apply MIL-SKDE in several applications of visual categorization, e.g., image and object categorization, and our algorithm performs superiorly comparing to other state-of-the-art methods. Furthermore, to address the complexity issue of MIL-SKDE, we propose MIL-SS (MIL with speed-up SKDE) to speed up the training process. Experiments shows that it has comparable performances to MIL-SKDE but is much more efficient in training stage.

Finally, we apply MIL-SS in a “bag-of-words” (BoW) system to learn the visual codebook for object categorization on a more comprehensive dataset. Our system consists of four steps: codebook generation, feature coding, feature pooling and classification. Unlike conventional BoW methods that learn codebook from the whole image areas, our method can learn codebook just from the areas of target objects, which significantly improves classification accuracy.

Author's publications for the Ph.D

Journal paper:

1. **Ruo Du**, Qiang Wu, Xiangjian He, and Jie Yang. "MIL-SKDE: Multiple-instance learning with supervised kernel density estimation". *Signal Processing*, Volume 93, Issue 6, June 2013, Pages 1471-1484

Conference papers:

2. **Ruo Du**, Qiang Wu, XiangJian HE, and Jie Yang. "Object categorisation based on a supervised mean shift algorithm". In *12th European Conference on Computer Vision (ECCV) Demos*, Part III, LNCS 7585. Springer, Heidelberg, 2012.
3. **Ruo Du**, Qiang Wu, XiangJian HE, and Jie Yang. "Multi-instance learning with an extended kernel density estimation for object categorisation". In *IEEE International Conference on Multimedia and Expo (ICME) Workshops*, pp.477-482, 9-13 July 2012.
4. Lin Wang, Xiangjian He, **Ruo Du**, Wenjing Jia, Qiang Wu, and Wei-Chang Yeh. "Facial expression recognition on hexagonal structure using lbp-based histogram variances". In *17th International MultiMedia Modeling Conference (MMM)*, pages 35 - 45, 2011.
5. **Ruo Du**, Sheng Wang, Qiang Wu, and Xiangjian He. "Learn concepts in multiple-instance learning with diverse density framework using supervised mean shift". In *Digital Image Computing: Techniques and Applications (DICTA) - Oral presentation*, pages 643-648, 2010.
6. Sheng Wang, **Ruo Du**, Qiang Wu, and Xiangjian He. "Adaptive stick-like features for human detection based on multi-scale feature fusion scheme". In *Digital Image Computing: Techniques and Applications (DICTA)*, pages 375-380, 2010.

-
7. **Ruo Du**, Qiang Wu, Xiangjian He, Wenjing Jia, and Daming Wei. “Facial expression recognition using histogram variances faces”. In *Workshop on Applications of Computer Vision (WACV)*, pages 1-7, 2009.

Contents

Contents	vii
List of Tables	xii
List of Figures	xiv
Nomenclature	xv
1 Introduction	1
1.1 Motivation	1
1.2 Facial expression recognition	3
1.2.1 Action unit based approaches	5
1.2.2 Emotion based approaches	6
1.2.3 Our approach for expression recognition	7
1.3 Object categorisation	8
1.3.1 Multiple-instance learning	10
1.3.2 Our approach for object categorisation	13
1.4 Contributions	13
1.5 Outline of this thesis	14
2 Related feature extraction and pattern recognition methods	16
2.1 Feature extraction	16
2.1.1 Haar-like features	17
2.1.2 Eigenface	20
2.1.3 Scale-invariant feature transform	22
2.1.4 Speeded up robust feature	27

2.1.5	Histogram of oriented gradients	28
2.2	Machine learning and pattern recognition	29
2.2.1	K-means	30
2.2.2	Kernel density estimation and mean shift	31
2.2.3	Adaboost	34
2.2.4	Support vector machines	36
2.2.5	Unsupervised and supervised topic models	38
2.2.5.1	Latent Dirichlet allocation (LDA)	38
2.2.5.2	Supervised latent Dirichlet allocation (sLDA)	41
2.2.5.3	Maximum entropy discrimination latent Dirichlet allocation (MedLDA)	43
2.3	Summary	46
3	Histogram Variances Faces for expression recognition	48
3.1	Histogram Variances Faces	49
3.1.1	Faces Alignment	49
3.1.2	Preprocessing and LBP texturising	51
3.1.3	Earth Mover's Distance for calculation of histogram variances	53
3.1.3.1	Earth Mover's Distance	53
3.1.3.2	Procedures of calculating histogram variances	55
3.1.3.3	Computing histograms of various-size blocks	56
3.2	Classifying HVF images using PCA+SVMs	57
3.2.1	PCA dimensionality reduction	57
3.2.2	SVMs training and recognition	57
3.3	Experiments	58
3.3.1	Dataset	58
3.3.2	Parameter selection for HVFs generation	60
3.3.3	Training and recognition	61
3.4	Discussion	63
3.5	Conclusion	64
4	Hexagonal Histogram Variances Faces for expression recognition	65
4.1	Hexagonal Histogram Variances Faces	66

4.1.1	Fiducial point detection and face alignment	66
4.1.2	Conversion from square structure to hexagonal structure .	66
4.1.3	Preprocessing and LBP texturising	67
4.1.4	Earth Mover's Distance (EMD)	68
4.1.5	Histogram variances	69
4.2	Classification	69
4.3	Experiments	70
4.3.1	Dataset	70
4.3.2	HHVFs generation	70
4.3.3	Training and recognition	71
4.4	Discussion	72
4.5	Conclusion	74
5	MIL-SKDE: Multiple-instance learning with supervised kernel density estimation	75
5.1	MIL-SKDE algorithm	75
5.1.1	Conventional kernel density estimation and mean shift . .	76
5.1.2	Supervised kernel density estimation	78
5.1.2.1	SKDE versus DDE	82
5.1.3	Supervised mean shift	83
5.1.3.1	Selecting starting points	85
5.1.3.2	Bandwidth estimation for supervised kernel density estimation	86
5.1.4	Algorithm summary	88
5.1.5	Classification	90
5.2	Experiments	91
5.2.1	Experiments on synthetic data	92
5.2.1.1	Mixture of positive and negative points	92
5.2.1.2	Unbalance of positive and negative points	92
5.2.1.3	Multiple concepts learning	93
5.2.2	Region-based image categorisation	94
5.2.2.1	Experiment setup	94
5.2.2.2	Image categorisation results	96

5.2.2.3	Sensitivity to labeling noise	97
5.2.3	Object categorisation	98
5.2.3.1	Experimental setup	99
5.2.3.2	Recognition results	101
5.2.3.3	Feature selection	103
5.3	Discussion	104
5.4	Conclusion	107
6	MIL-SS: Multiple-instance learning with a speed-up supervised kernel density estimation	108
6.1	MIL-SS algorithm	109
6.1.1	Revisit supervised kernel density estimation and mode seeking	109
6.1.2	Instance selection process	111
6.1.3	Bandwidth estimation	113
6.1.4	MIL-SS algorithm summary	113
6.1.5	Classification	114
6.2	Experiments	114
6.2.1	Region-based image categorisation	114
6.2.2	Object categorisation	115
6.2.3	Sensitivity to labeling noise	118
6.2.4	Regions of interest detection	120
6.3	MIL-SS for bag-of-words model	120
6.3.1	Bag-of-words model	122
6.3.1.1	Feature extraction	124
6.3.1.2	Codebook generation	124
6.3.1.3	Feature coding and pooling	125
6.3.1.4	Classification	125
6.3.2	Experiments on SIVAL dataset	126
6.4	Conclusion	126
7	Conclusions and future work	128
7.1	Conclusions	128

CONTENTS

7.2 Future work	129
References	131

List of Tables

3.1	Recognition rates of happy and surprise HVFs.	61
3.2	A recent investigation of facial expression recognition by human .	61
3.3	Recognition rates of happy and surprise versus other sorts of HVFs.	62
3.4	Recognition rates of anger, disgust, surprise and sadness HVFs. .	62
3.5	Recognition rates of all sorts of HVFs.	63
4.1	Recognition rates of happy and surprise HHVFs.	71
4.2	A recent investigation of facial expression recognition by human .	72
4.3	Recognition rates of anger, disgust, surprise and sadness HHVFs.	72
4.4	Recognition rates of all sorts of HHVFs	73
4.5	Recognition rates between HVFs and HHVFs. The last row is the average recognition rates of six categories. In our experiments, HHVFs slightly outperforms HVFs.	73
5.1	Learning for multiple concepts.	93
5.2	Comparison of image categorisation accuracy rates for MIL-SKDE and other methods.	97
5.3	Confusion matrices of object categorisation on Caltech-4 and SIVAL dataset respectively.	103
5.4	The recognition rates for MIL-SKDE, DD-SVM, MILIS and MILES.	104
6.1	Comparison of image categorisation accuracy rates for MIL-SKDE and other methods.	116
6.2	The recognition rates for MIL-SS, MIL-SKDE, DD-SVM, MILIS and MILES.	117

LIST OF TABLES

6.3	The average training scales and time uses of MIL-SS and MIL-SKDE for object categorisation.	118
6.4	Classification accuracy comparisons among different methods over 30 runs on the SIVAL.	127

List of Figures

1.1	The training process for visual categorisation.	2
1.2	The testing process for visual categorisation.	2
1.3	Examples of some Action Units extracted from Cohn-Kanade database	4
1.4	Examples of some combinations of Action Units.	4
1.5	The task of object-based visual categorisation.	8
1.6	Examples of salient parts.	9
2.1	Examples of Haar-like features.	18
2.2	Computing the sum of a rectangular area using integral image. . .	19
2.3	Typical Haar-like features for face detection [68].	19
2.4	Visualisation of the eigenface approach	20
2.5	Generation of Difference of Gaussians (DoG)	23
2.6	SIFT keypoint candidature detection.	24
2.7	Generation of SIFT descriptor.	28
2.8	Maximum-margin hyperplane obtained by an SVM	36
2.9	LDA graphical model.	40
2.10	Graphical model representation of the variational distribution q . .	41
2.11	Supervised topic models (sLDA).	42
2.12	Graphical model of MedLDA.	44
3.1	Procedures of generating a HVF image.	50
3.2	An example of computing LBP in a 3×3 neighborhood	52
3.3	Examples of HVF images	56
3.4	Some example sequences in the Cohn-Kanade database.	59

LIST OF FIGURES

4.1	A 9x8 square structure and a constructed 7x8 hexagonal structure.	67
4.2	An example of computing LBP in a 3×3 neighborhood on square structure.	68
4.3	An example of computing HLBP in a 7-pixel hexagonal cluster. .	68
4.4	Examples of HHVF images	70
5.1	The properties of concepts in a feature space.	80
5.2	One of the iterations of the supervised mean shift	85
5.3	Kernel density estimate with different bandwidths.	88
5.4	Supervised mean shift on simulated data for mixture of positive and negative data.	93
5.5	Supervised mean shift on simulated data for local mode displacement.	94
5.6	Supervised mean shift on simulated data for unbalance data. . . .	95
5.7	Image samples from the COREL image database for region-based image categorisation.	96
5.8	Comparison of sensitivity to labeling noise among MIL algorithms.	98
5.9	Some image samples from the Caltech-4 dataset for object categorisation.	100
5.10	Some image samples from the SIVAL dataset.	100
5.11	An example of instances detection and matching using SIFT. . . .	101
5.12	A figure shows bandwidth calculation on SIVAL dataset using Algorithm 3.	102
5.13	Several sample images that are recognised as containing a target object.	105
6.1	Remained training data after instance selection process.	112
6.2	Comparisons of sensitivity to labeling noise among different methods.	119
6.3	The regions of interest detected by MIL-SS.	121
6.4	General procedures for a bag-of-words model.	123